

## SURVEYING THE E-JOURNAL PRESERVATION LANDSCAPE

by Anne R. Kenney, Associate University Librarian,  
Instruction, Research, and Information Services,  
Cornell University Library

“Digital preservation represents one of the grand challenges facing higher education,” wrote a working group of influential academic administrators and librarians who participated in a special meeting convened at The Andrew W. Mellon Foundation in September 2005.<sup>1</sup> Their statement, titled “Urgent Action Needed to Preserve Scholarly Electronic Journals,” signaled an intensity of broad concern and called the educational community to action. The statement underscored the fact that preserving electronic publications has become a critical matter as the mass of e-publication increases and our user communities have begun to depend on electronic publications as they used to rely on paper.

The Council on Library and Information Resources (CLIR) and ARL believe that libraries require a better understanding of the emerging strategies and options for ensuring long-term access to the born-digital scholarly literature in order to determine their best course of action. The two organizations agreed that a framework could be developed to describe preservation strategies for peer-reviewed journal literature and to assess the scope and range, potential, and vulnerabilities of such strategies. This framework could be used to survey the most promising preservation programs to reveal opportunities for investment.

The Scholarly Communication Steering Committee of ARL, with a long history of expressing the concerns of the leaders in the research library community, has sought the collaboration and support of CLIR to develop a landscape analysis for preserving e-journals. This article is a preliminary report about that project.

With its history of undertaking and managing rigorous research projects of this nature, CLIR accepted the commission from ARL and contracted with the Cornell University Library Research & Assessment Services department for the landscape analysis. The research is a team effort, involving the work of Ellie Buckley (Digital Research Specialist), Richard Entlich (Digital Projects Librarian), Peter Hirtle (Technology Coordinator and Intellectual Property Officer), Nancy McGovern (Department Director and Digital Preservation Officer), and Anne Kenney.

The study’s focus is the “who, what, when, where, why, and how” of significant preservation programs operated by not-for-profit organizations in the domain of peer-reviewed journal literature published in digital form. At the center of this work are 10 initiatives that acknowledge preservation responsibility for e-journal

archiving; the team will also identify other promising efforts in planning or pilot stages.

### Background

We know that scholars, publishers, libraries, consortia, and other organizations have stirred into action and we have seen a flurry of recent initiatives:

- publishers collaborating with cultural institutions to provide dark archives for their back files;
- in several countries, passage of legal deposit laws that include rights to preserve electronic journal content;
- the National Institutes of Health’s (NIH) decision to create an archive of accessible, government-funded research publications and the corresponding protests from commercial and not-for-profit publishers and societies;
- national libraries establishing or financially supporting e-journal archiving programs;
- launch of third-party and consortial efforts that focus on e-journals;
- development of a draft *Audit Checklist for Certifying Digital Repositories* by the Research Libraries Group (RLG) and the National Archives and Records Administration (NARA);<sup>2</sup> and
- road testing of the RLG-NARA certification requirements by the Center for Research Libraries in several digital repositories, with a heavy focus on e-journal preservation and an eagerly awaited report on the results due this fall.

The “Urgent Action” statement argued for a four-pronged approach. First, the community should recognize that preservation of e-journals is a “kind of insurance, and is not in and of itself a form of access.” Second, preservation archives should provide a minimal set of well-defined services. Third, libraries must invest in a qualified archiving solution. Fourth, libraries must demand archival deposit by publishers as part of their licensing agreements. Some organizations have already endorsed or supported the manifesto, including Association of College and Research Libraries (ACRL), Association for Library Collections and Technical Services (ALCTS), ARL, Consortium of Academic and Research Libraries in Illinois (CARLI), International Coalition of Library Consortia (ICOLC), Medical Library Association, and NorthEast Research Library Consortium (NERL). Other groups are considering endorsement as well. ACRL, in particular, expects to develop “guidelines and effective practices for academic libraries in this area.”

### Ten E-Journal Archiving Initiatives

The 10 e-journal archiving initiatives that the study team

# CURRENT ISSUES

*Continued*

has identified and intends to evaluate further are briefly described below.

## **Koninklijke Bibliotheek e-Depot**

<http://www.kb.nl/dnp/e-depot/e-depot-en.html>

As the national deposit library for the Netherlands, the Koninklijke Bibliotheek is responsible for preserving and providing long-term access to Dutch electronic publications. Consequently, it has developed e-Depot: a fully automated system, dedicated to long-term storage and large-scale archiving. It is primarily intended for archiving publications by Dutch publishers, and currently offers digital archiving services for nine major publishers, including some outside the Netherlands.

## **KOPAL**

<http://kopal.langzeitarchivierung.de/>

Funded by the German Federal Ministry of Education and Research, KOPAL is a cooperative project begun in July 2004. Its goal is to develop an innovative technical solution to the problem of how to keep digital documents accessible over time. Project partners, Die Deutsche Bibliothek and the Lower Saxon State and University Library (SUB Göttingen), are storing a variety of digital materials in a repository based on DIAS, the Digital Information and Archiving System, developed by IBM and the Koninklijke Bibliotheek, in The Hague. The Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWDG) is in charge of the archive's technical operation, with software support provided by IBM Deutschland GmbH. In the future, KOPAL intends to help other institutions keep their data available on a long-term basis.

## **Los Alamos National Laboratory Library**

<http://library.lanl.gov/lww/>

The Research Library at Los Alamos National Laboratory (LANL) has been locally loading licensed back files from a variety of commercial and society publishers since 1995. The library provides the content to LANL staff and others (universities and the Department of Energy) that have licensed it on a cost-recovery basis. LANL's commitment to maintaining its back files depends on availability of funding and on whether alternative options emerge for access to the content.

## **LOCKSS Alliance/CLOCKSS**

<http://www.lockss.org/clockss/>

The Lots of Copies Keeps Stuff Safe (LOCKSS) program based at Stanford launched the beta version of its open source software between 2000 and 2002. LOCKSS intended the software to allow libraries to collect, store, preserve, and provide access to their own local copy of authorized content they purchase. More than 80 institutions in over 20 countries are using the LOCKSS software to capture content. More than 50 publishers,

largely not-for-profit or open access, are participating in the LOCKSS program. In 2005, the LOCKSS Alliance was launched as a membership organization to introduce governance and to address sustainability issues. The Community LOCKSS (CLOCKSS) initiative is a recent addition to the LOCKSS program, bringing together six libraries and nine publishers to establish a large dark archive for e-journals.

## **National Library of Australia**

<http://pandora.nla.gov.au/>

The National Library of Australia selects e-journals from its Australian Journals Online database for preservation in PANDORA (Preserving and Accessing Networked Documentary Resources of Australia), which was established in 1996. E-journals represent one of six categories of online publications included in PANDORA, which lists a total of more than 11,000 titles for all six categories. The first version of the PANDORA Archiving System (PANDAS) was released in 2001.

## **OCLC ECO**

<http://www.oclc.org/electroniccollections/>

OCLC's Electronic Collections Online (ECO) is an electronic journals service that offers Web access to a collection of more than 5,000 titles in a wide range of subject areas, from over 70 publishers of academic and professional journals. OCLC has negotiated with publishers to secure subscribers' perpetual rights to journal content. In addition, OCLC has reserved the right to migrate journal backfiles to new data formats as they become available.

## **OhioLINK**

<http://www.ohiolink.edu/>

The Ohio Library and Information Network (OhioLINK) is a consortium of Ohio's college and university libraries, comprising 85 institutions of higher education and the State Library of Ohio. OhioLINK's electronic services include a multi-publisher Electronic Journal Center (EJC), launched in 1998, which contains more than 6,400 scholarly journal titles from more than 80 publishers across a wide range of disciplines. OhioLINK has declared its intention to maintain the EJC content as a permanent archive and has acquired perpetual archival rights in its licenses from all publishers but one.

## **Ontario Scholars Portal**

<http://www.scholarsportal.info/>

The Ontario Scholars Portal serves all 20 university libraries in the Ontario Council of University Libraries (OCUL). The portal includes 7,500 e-journals from about 20 publishers, and metadata for the content of an additional three publishers. The primary purpose of the portal is access, but OCUL has made an explicit commitment to the long-term preservation of the

e-journal content it loads locally. The initiative began with grant funding and became self-funded through tiered membership fees on January 1, 2006.

### **Portico**

<http://www.portico.org/>

Portico is a third-party electronic archiving service for e-journals that has received support from The Andrew W. Mellon Foundation, Ithaka, Library of Congress, and JSTOR. At present, seven publishers have agreed to participate in Portico. Publishers and libraries are both asked to support the effort through annual contributions. Recently announced library fees, ranging from \$1,500 to \$24,000 per annum, are based on the total library materials expenditures for an individual institution.

### **PubMed Central**

<http://www.pubmedcentral.nih.gov/>

Launched in February 2000, PubMed Central (PMC) is the NIH's free digital archive of biomedical and life sciences journal literature, run by the National Center for Biotechnology Information of the National Library of Medicine. PMC currently encompasses approximately 220 titles from 40 publishers. PMC prefers that participating titles submit all content but will accept, at minimum, the primary research content. PMC allows publishers to delay deposit by a year or more after initial publication. It retains perpetual rights to archive all submitted materials and has made a commitment to maintain the long-term integrity and accuracy of the contents of the archive.

### **Designing the Study**

The Cornell team began the study by developing a sense of the key e-preservation areas that library decision makers are likely to consider as they assess preservation strategies. Feedback from directors of member libraries of the Center for Research Libraries after sessions held at the 2006 American Library Association Midwinter Meeting was particularly helpful in framing the initial list.<sup>3</sup> The team canvassed library directors to understand their greatest needs and the constraints involved in making these judgments. Their concerns will guide the design of a structured survey format to be used in appraising 10 ten e-journal preservation efforts described above.

Telephone interviews explored a set of six key e-preservation areas that library decision makers are likely to consider. The team contacted 15 library directors across North America, representing a range of public and private institutions of various sizes as well as consortia.

The six key areas are:

1. Library motivation (*Why should we be concerned about or invest in this?*)

2. Content coverage (*Are current approaches covering the subject areas, titles, and journal components we're most interested in?*)
3. Access (*What will we gain access to? When and under what conditions?*)
4. Program viability (*What evidence is there that these efforts are sufficiently well-governed and financed to last?*)
5. Library responsibilities/resource requirements (*What will this cost our library in staff time, expertise, financial commitment? Would our support result in any cost savings to the library?*)
6. Technical approach (*How do we judge whether the approach is rigorous enough to meet its preservation objectives?*)

### **Next Steps**

The study team will now take the directors' concerns and develop a survey, which will be used to interview the principals at the 10 e-journal archiving initiatives. The survey will explore technical functions, such as ingestion, data validation, storage management, and preservation planning; business practices, funding models, and organizational viability; content and coverage; access considerations, including timing and level of effort; trigger events (i.e., what has to happen to open the preservation archive for use?); publisher relations; and library responsibilities.

After the reviews are completed in early April, the team will analyze the data to provide a neutral structure for contrast and comparison, rather than evaluation or measurement against a standard. The goal is to present information comprehensibly, as a basis for informed decision making by library directors. This snapshot and the underlying analysis will continue to be useful as new options become available.

The study will be previewed at the ARL Membership Meeting in Ottawa, Ontario, in mid-May, and the final report will be published by CLIR by mid-August. As the investigation continues, the Cornell team welcomes observations and suggestions from others.

To offer feedback or obtain more information about the study, send e-mail to [irisadmin@cornell.edu](mailto:irisadmin@cornell.edu).

<sup>1</sup> "Urgent Action Needed to Preserve Scholarly Electronic Journals," October 15, 2005, <http://www.diglib.org/pubs/waters051015.htm>.

<sup>2</sup> RLG and NARA, *Audit Checklist for Certifying Digital Repositories*, [http://www.rlg.org/en/page.php?Page\\_ID=20769](http://www.rlg.org/en/page.php?Page_ID=20769).

<sup>3</sup> "Digital Repositories: Some Concerns and Interests Voiced in the CRL Directors' Conversation January 21–22, 2006 [at ALA Midwinter]," as distributed on the CRL member directors electronic discussion list, February 03, 2006, by Bernard F. Reilly, President of CRL. See also "Digital Archives and Repositories Update," *FOCUS* 25, no. 2 (Winter 2005–06), <http://www.crl.edu/PDF/pdfFocus/Winter2005-06.pdf>